DOCUMENT RESUME

ED 459 290 UD 034 589

AUTHOR	Darling-Hammond, Linda
TITLE	Transforming Urban Public Schools: The Role of Standards and Accountability.
PUB DATE	2000-00-00
NOTE	55p.
PUB TYPE	Opinion Papers (120)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Academic Achievement; *Academic Standards; *Accountability; Educational Change; *Educational Quality; Elementary Secondary Education; Equal Education; High Stakes Tests; Minority Group Children; Public Schools; Standardized Tests; Student Evaluation; Teacher Competencies; *Urban Schools

ABSTRACT

This paper examines how urban school districts that have substantially improved student performance emphasize improvement of education guided by rigorous standards for teachers rather than high-stakes testing for students. States and districts that rely on test-based accountability emphasizing sanctions for students and teachers often produce greater failure rather than success for educationally vulnerable students. The paper reviews research on various approaches to accountability and highlights successful reforms in urban settings that emphasize the use of standards for teaching and learning to guide investments in better prepared teachers, higher quality teaching, more performance-oriented curriculum and assessment, better designed schools, more equitable and effective resource allocations, and more diagnostic supports for student learning. It argues for a broader conception of accountability that focuses on whether policymakers' and practitioners' actions, in fact, produce better quality education and higher levels of learning for a greater share of students. It suggests that genuine accountability is achieved when school system policies and operating practices work both to provide good education and to correct problems as they occur. It concludes that raising standards for students so they learn what they need to know requires raising standards for the system. (Contains 69 references.) (SM)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 459 290

Transforming Urban Public Schools: The Role of Standards and Accountability

Linda Darling-Hammond

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

L. Darling-Hammond

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

BEST COPY AVAILABLE

UD034589

Transforming Urban Public Schools: The Role of Standards and Accountability

Linda Darling-Hammond

The education reform movement in the United States has begun to focus increasingly on the development of new standards for students: Virtually all states have begun the process of creating standards for student learning, new curriculum frameworks to guide instruction, and new assessments to test students' knowledge. President Clinton and presidential candidates Gore and Bush all proposed testing as a centerpiece of accountability reforms, and many school districts across the country are weighing in with their own versions of standards-based reform, including new curricula, testing systems, accountability mechanisms, and promotion or graduation requirements.

The rhetoric of these reforms is appealing. Students cannot succeed in meeting the demands of the new economy if they do not encounter much more challenging work in school, many argue, and schools cannot be stimulated to improve unless the real accomplishments -- or deficits -- of their students are raised to public attention. There is certainly merit to these arguments. But will standards and tests improve schools or create educational opportunities where they do not now exist?

Some proponents of standards-based reforms have envisioned that standards that express what students should know and be able to do would spur other reforms that mobilize more resources for student learning, including high quality curriculum frameworks, materials, and assessments tied to the standards; more widely available course offerings that reflect this high quality curriculum; more intensive teacher preparation and professional development guided by related standards for teaching; more equalized resources for schools; and more readily available

safety nets for educationally needy students (O'Day and Smith, 1993). For others, the notions of standards and 'accountability' have become synonymous with mandates for student testing which may have little connection to policy initiatives that directly address the quality of teaching, the allocation of resources, or the nature of schooling.

In addition to these differences, distinct change theories have emerged around the idea of standards-based reform. Some argue that standards for learning and teaching should be used primarily to inform investments and curricular changes that will strengthen schools. They see the major problem as a need for teacher, school, and system learning about more effective practice combined with more equal and better-targeted resource allocation. Others argue that standards can motivate change only if they are used to apply sanctions to those who fail to meet them. They see the major problem as a lack of effort and focus on the part of educators and students.

Policy makers who endorse the latter view have emphasized "high stakes" testing – that is, the use of scores on achievement tests to make decisions that have important consequences for examinees and others – as a primary strategy to promote accountability. Some high stakes decisions affect students, such as the use of test scores for promotion, tracking and graduation. Others affect teachers and principals when scores are used to determine merit pay or potential dismissal. Still others affect schools, as when schools are awarded recognition or extra funds when scores increase or are put into intervention status or threatened with loss of registration when scores are low. Some policies take into account differences in the initial performance of students and in the many non-school factors that can affect achievement. Some do not, holding schools to similar standards despite dissimilar student populations and resources.

Many questions arise from this policy strategy. Will investments in better teaching, curriculum, and schooling follow the press for new standards? Or will standards built upon a

foundation of continued inequality in education simply support tests that more effectively certify student failure and reduce access to future education and employment? In states where high-stakes tests have been imposed without attention to existing educational inequality and access to the curriculum and teaching needed to meet the standards, a new generation of equity lawsuits has emerged. Litigation in California, Florida, New York, and elsewhere has followed on the heels of recently successful "adequacy" lawsuits in Alabama and New Jersey.

A body of research has emerged that suggests that accountability strategies that have relied primarily on high-stakes testing programs have not always improved the quality of instruction or the outcomes for educationally needy students. A number of studies have found that high-stakes tests can narrow the curriculum, pushing instruction toward lower order cognitive skills, and can distort scores (Klein et al., 2000; Koretz and Barron, 1998; Koretz et al., 1991; Linn, 2000; Linn, Graue, and Sanders, 1990; Stetcher et al., 1998). In addition, there is evidence that high-stakes tests that reward or sanction schools based on average student scores have created incentives for pushing low-scorers into special education, holding them back in the grades, and encouraging them to drop out so that schools' average scores will look better (Shepard and Smith, 1988; Koretz, 1988; Smith et al., 1986; Darling-Hammond, 1991, 1992; Allington and McGill-Franzen, 1992). School incentives tied to test scores have undermined some districts' efforts to create and sustain more inclusive and integrated student populations, especially in emerging schools of choice, as test-based sanctions punish schools for accepting and keeping students with high levels of special needs and reward them for keeping such students out of their programs through selective admissions, transfer, and even push out policies (Smith et al., 1986; Haney, 2000).

In this paper, I show that urban districts that have substantially improved their students' performance have focused on the improvement of teaching guided by rigorous standards for teachers rather than on high-stakes testing for students. States and districts that have relied primarily on test-based accountability emphasizing sanctions for students and teachers have often produced greater failure rather than greater success for their most educationally vulnerable students. I review research on these approaches to "accountability," and I provide examples of more successful reforms in urban settings that have emphasized the use of standards for teaching and learning to guide investments in better prepared teachers, higher quality teaching, more performance-oriented curriculum and assessment, better designed schools, more equitable and effective resource allocations, and more diagnostic supports for student learning.

In the course of this paper, I also argue for a broader conception of accountability that focuses on whether the actions undertaken by policymakers and practitioners in fact produce better quality education and higher levels of learning for a greater share of students. Genuine accountability, I suggest, is achieved when a school system's policies and operating practices work both to provide good education and to correct problems as they occur. Assessment data are helpful in this regard to the extent that they provide relevant, valid, timely, and useful information about how individual students are doing and how schools are serving them. But this kind of data is only a tiny part of the total process. Indicators such as test scores or dropout rates are information for the accountability system; they are not the system itself. Accountability occurs only when a useful set of processes exists for interpreting and acting on the information in educationally productive ways.

This definition of accountability suggests that we should gauge policy strategies on the basis of whether and for whom they provide greater assurance of high quality teaching and

learning. We should ask who is helped and who is harmed by policies that are offered under the name of accountability. Do "accountability" systems heighten the probability that good practices will occur for students and reduce the likelihood that harmful practices will occur? And do they provide self-correctives in the system to identify, diagnose, and changes courses of action that are harmful or ineffective?

Based on this definition and on evidence about successful urban school reforms, I propose the following policy recommendations:

Accountability that supports learning for a wide range of students should focus on: 1) the quality of teachers and teaching available to individuals and groups of students, 2) the availability of high quality curriculum, 3) the creation of informative and appropriately used assessments that reflect student learning standards, and 4) the design of schools so that they can focus on student needs and in-depth learning.

In particular, accountable school systems will:

- **Develop and use criterion-referenced standards for student learning that reflect core concepts and important skills relevant to success in today's society and economy.**
- **Use standards-based performance assessments of student achievement as diagnostics to guide improved teaching and needed supports, not as the basis for punishing students.**
- **Use standards for teaching to hire and support teachers in all schools and classrooms who are well-prepared to help diverse students meet student learning standards.**
- **Offer professional learning opportunities that build teachers' capacity to teach effectively, use assessments to inform their teaching, and meet the needs of struggling students.**
- **Assure that curriculum materials and course offerings responsive to student learning standards are available in all schools and are accessible to all students.**
- **Develop school organizations that enable teachers to teach for in-depth learning and to know their students well.**

- **Create strategies for school accountability that continually examine the appropriateness and adequacy of students' learning opportunities and create levers and supports for school change.**

Ultimately, raising standards for students so that they learn what they need to know requires raising standards for all parts of the educational system, so that it provides the kinds of teaching and school settings students need in order to learn. In what follows I describe several possible approaches to accountability. I discuss the problems with many current approaches to bureaucratic accountability in urban school systems and the possibilities revealed by systems that have taken a comprehensive approach to accountability that begins with improving the quality and equality of teaching and schooling as a means for upgrading learning.

Types of Educational Accountability

In education, as in other enterprises in our society, at least five types of accountability mechanisms exist alongside each other:

- political accountability - legislators and school board members, for example, must regularly stand for election.
- legal accountability - citizens can ask the courts to hear complaints about the public schools' violation of laws, say, regarding desegregation or equal educational opportunity.
- bureaucratic accountability - district and state education offices promulgate rules and regulations intended to ensure that schooling takes place according to set standards.
- professional accountability - teachers and other school staff must acquire specialized knowledge, pass certification exams, and uphold professional standards of practice.

- market accountability - parents and students may choose the courses or schools they believe are most appropriate. They may also be involved in other, more direct means of participating in school decision-making.

All of these accountability mechanisms have their strengths and weaknesses, and each is more or less appropriate for certain goals. Political mechanisms can help establish general policy directions, but they do not allow citizens to judge each decision by elected officials, and they do not necessarily secure the rights of minorities. Legal mechanisms are useful in establishing and defending rights, but not everything is subject to court action and not all citizens have access to the courts. Bureaucratic mechanisms are appropriate when standard procedures will produce desired outcomes, but they can be counterproductive when clients have unique and needs that require differential responses by those who must make non-routine decisions. Professional mechanisms are important when services require complex knowledge and decision making to meet clients' individual needs, but they do not always take competing public goals (e.g. cost containment) into account. Market mechanisms are helpful when consumer preferences vary widely and the state has no direct interest in controlling choice, but they do not ensure that all citizens will have access to services of a given quality.

Because of these limits, no single form of accountability operates alone in any major areas of public life. The choices of accountability tools -- and the balance among different forms of accountability -- are constantly shifting as problems emerge, as social goals change, and as new circumstances arise.

In public education, the power of electoral accountability, exemplified in the authority of school boards, has waxed and waned over the past 20 years relative to that of appointed officials. During the same period, legal forms of accountability have grown, as court cases have been used

to settle educational policy; and bureaucratic forms of accountability have expanded through increased policy-making at the district and state levels. Recently, market accountability, the least used form, has been expanded somewhat through magnet schools, charters, and other schools of choice. Finally, based on an expanding knowledge base and efforts to create more meaningful standards of practice, professional accountability has gained currency as a way to improve teaching.

In most urban public school systems, legal and bureaucratic accountability strategies have predominated over the last twenty or more years. These have especially focused on attempts to manage schooling through standardized educational procedures, prescribed curriculum and texts, and test-based accountability strategies, often tied to tracking and grouping decisions that are meant to determine the programs students will receive. Few have experimented with market accountability. Most notable among them are New York City, which has launched and maintained more than 150 small schools of choice in the last decade to add to the many dozens that existed before that time, and Cambridge, Massachusetts which has had a system of choice-based schools for more than ten years. (Limited forays into vouchers for low-income students in Milwaukee, Wisconsin and Cleveland, Ohio have been extremely small in scale. Since the vouchers were small and limited in scope, and since most private schools were not willing to accept either the vouchers or the students who came with them, it is not plausible to describe these as well-developed market experiments.) Finally, a very few urban districts have launched well-developed professional accountability strategies tied to standards for teaching as well as student learning. New York City's District #2, New Haven, California, and several cities in Connecticut, a state that launched a highly successful statewide reform focused on teaching quality are among these, and are described later.

Standards as Assessment: Attempts to Create Accountability through High-Stakes Testing

Since the mid-1800s, a number of urban districts have periodically used student test scores to allocate rewards or sanctions to schools or teachers. Many have approached standards-based reform through this familiar strategy, alleging that they have implemented new standards even when the tests are not aligned to their new curriculum frameworks and when there is little evidence that students are assured of receiving qualified teachers, curriculum aligned with the standards, or schools organized to support them. Several cities' "standards-based reform strategies" that have emphasized using test scores as the basis for promoting students from grade to grade, allocating program opportunities, and making graduation decisions have received a great deal of publicity in the mid- to late- 1990s as "new" reforms; however, they replicate policies that have come and gone many times over the last century or more.

In contrast to schools in most European and Asian countries, U.S. schools have a long tradition of retaining students in a grade if they seem not to be succeeding at school. It has been estimated that the United States has an overall retention rate of 15-20 percent of its students annually (most of them at-risk students in central cities), placing U.S. public schools on a par with countries like Haiti or Sierra Leone and in stark contrast with European nations that bar grade retention or countries like Japan, which has less than a 1 percent rate of grade retention (Smith and Shepard, 1987). Most recently, during the early 1980s, grade retentions increased as school districts instituted policies that linked standardized test scores to student promotion and placement decisions. However, many of these policies failed and were repealed by the late 1980s, only to be re-instated less than a decade later.

For example, New York City experienced many of the problems associated with grade retention when the "Promotional Gates Program" was put in place in elementary and junior high

schools during the early 1980s. At that time, gateways in grades four and eight were created through which students could pass only if they demonstrated a specified level of performance on the standardized citywide reading and mathematics tests. Students who did not meet the minimum standards were retained, sometimes repeatedly, unless or until they were able to achieve the necessary score on the tests. The actual outcomes of this program however, did not meet the creators' original goals of strengthening students' academic performance and heightening their academic self-confidence. Instead, the program created cohorts of students who had been retained repeatedly; sometimes they had been held back for so long that their advanced physical size and age caused a host of problems, including increased misbehavior and decreased achievement for both the retained students and others in their classrooms.

The negative consequences of having 12-year-olds in the fourth grade and 16-year-olds who were still in junior high school ultimately meant that students eventually had to be moved on to the next grade even if their academic achievement had not improved. Frequently, however, the damage was done: The students retained had lower achievement, greater incidences of disciplinary difficulties, and higher dropout rates while their teachers' jobs were made ever more unmanageable. Eventually, in the face of local and nationwide evidence about the failures of such an approach, the program was ended by Chancellor Fernandez in the late 1980s (Gampert and Opperman, 1988).

A decade later, with no sense of irony or institutional memory, the New York Times reported in September, 1999 that 21,000 students would be held back under the City's "new" policy to end social promotion (Wasserman, 1999). Two weeks later the newspaper reported that the social promotion policy was in disarray as two-thirds of the 35,000 students forced to take summer school still did not pass the tests and, further, that 4,500 students' test scores had been

misreported and as many as 3,000 had been forced to take summer school by mistake (Hartocollis, 1999). Similar news headlines appeared in Los Angeles, where a policy to "end social promotion" resulted in more than 10,000 students being threatened with grade retention, only to find that the schools could not accurately identify who had passed or failed and could not find qualified teachers to teach the summer school programs that were supposed, miraculously, to catch these students up.

These outcomes have been replicated in more recent test-based promotion and graduation reforms. For example, the much publicized Chicago effort, which sought to end social promotion by requiring test passage at grades 3, 6, and 8 appears to have failed to improve the learning of the thousands of students it retained. In the first two years under the policy, more than one-third of 3rd, 6th, and 8th graders failed to meet the promotional test cutoffs by the end of the school year. Despite the fact that there were large-scale waivers for students with limited English proficiency and special education students, more than 20,000 students were retained in grade in 1997 and 1998, during the first two years of the program. An evaluation by Consortium on Chicago School Research concluded that:

> Retained students did not do better than previously socially promoted students. The progress among retained third graders was most troubling. Over the two years between the end of second grade and the end of the second time through third grade, the average ITBS reading scores of these students increased only 1.2 GEs (grade equivalents) compared to 1.5 GEs for students with similar test scores who had been promoted prior to the policy. Also troubling is that one-year dropout rates among eighth graders with low skills are higher under this policy…. In short, Chicago has not solved the problem of poor performance among those who do not meet the minimum test cutoffs and are retained. Both the history of prior attempts to redress poor performance with retention and previous research would clearly have predicted this finding. Few studies of retention have found positive impacts, and most suggest that retained students do not better than socially promoted students. The CPS policy now highlights a group of students who are facing significant barriers to learning and are falling farther and farther behind (Roderick, Bryk, Jacob, Easton, & Allensworth, 1999, pp. 55-56).

These findings are not surprising. In recent years a substantial body of research has demonstrated that retaining students does not help them catch up with peers and succeed in school; meanwhile, it contributes to academic failure and behavioral difficulties. Studies comparing the learning gains of students who were retained with those of academically comparable students who were promoted have found that retained students actually achieve less than their comparable peers who move on through the grades. Students do not appear to benefit academically from grade retention regardless of the grade level or the student's initial achievement level (Baenen, 1988; Gampert and Opperman, 1988; Hess, 1987; Holmes and Matthews, 1984; Illinois Fair Schools Coalition, 1985; Labaree, 1984; Medway, 1985; Meisels, 1992; Oakes and Lipton, 1990; Ostrowski, 1987; Safer, 1986; Shepard and Smith, 1988; Smith and Shepard, 1987; Walker and Madhere, 1987). Shephard and Smith (1986) conclude in their review of research: "Contrary to popular beliefs, repeating a grade does not help students gain ground academically and has a negative impact on social adjustment and self-esteem" (p. 86).

When students who were retained in grade are compared with students of equal achievement levels who were promoted, the retained students consistently suffer poorer self-concepts, have more problems of social adjustment, and express more negative attitudes toward school at the end of the period of retention than do similar students who are promoted (Eads, 1990; Holmes and Matthews, 1984; Illinois Fair Schools Coalition, 1985; Natale, 1991; Shepard and Smith, 1988; Walker and Madhere, 1987).

In addition, many studies over the past two decades have found that grade retention increases dropout rates (Hess, 1986; Hess, Ells, Prindle, Liffman, and Kaplan, 1987; Massachusetts Advocacy Center, 1988; Safer, 1986b; Smith and Shepard, 1987). Initial retentions frequently result in multiple retentions. A single grade retention increases the

likelihood of a student dropping out by 40 to 50 percent. A second retention increases this risk by 90 percent (Carnegie Council on Adolescent Development, 1987; Massachusetts Advocacy Center, 1988; Wehlage, Rutter, Smith, Lesko, and Fernandez, 1990).

It is not surprising that these outcomes of grade retention policies appear to be consistent. The concept of holding students back is a crude remedy for educational problems derived from the factory assembly line model of schooling developed during the early years of the twentieth century: The assumption was that a sequenced set of procedures would be implemented as a child moved along the conveyor belt from 1st to 12th grade. If a particular set of procedures didn't "take," the procedures should be repeated until the child was properly "processed." This approach to educational correction does not address questions of whether the child was appropriately taught the first time; whether different strategies might be needed to support learning; or whether the educational environment itself is lacking.

The premise of grade retention as a solution for poor performance is that the problem, if there is one, resides in the child, rather than in the teaching and schooling setting he or she has encountered. Rather than looking carefully at classroom or school practices when students are not achieving, schools send students back to repeat the same experience over again. Very little is done to ensure that the experience will be either higher quality or more appropriate for the individual needs of the child. There is no accountability for the quality of the educational experience students receive.

This is particularly troubling given mounting evidence that children receive dramatically unequal access to high-quality curriculum and teaching, and that these differentials are strongly related to their achievement (see Darling-Hammond, 1997, for a review). Despite the rhetoric of American equality, the school experiences of students of color in the United States continue to be

substantially separate and unequal. Nearly two-thirds of "minority" students attend predominantly minority schools, and one-third of Black students attend intensely segregated schools (i.e. 90% or more minority enrollment), most of which are in central cities. Currently, about 2/3 of all students in central city schools are Black or Hispanic (NCES, 1997a). This concentration facilitates inequality. Not only do funding systems and tax policies leave most urban districts with fewer resources than their suburban neighbors, but schools with high concentrations of "minority" students receive fewer resources than other schools within these districts. And tracking systems exacerbate these inequalities by segregating many "minority" students within schools, allocating still fewer educational opportunities to them at the classroom level.

In their review of resource allocation studies, MacPhail-Wilcox & King (1986) summarize the resulting situation as follows:

> School expenditure levels correlate positively with student socioeconomic status and negatively with educational need when school size and grade level are controlled statistically.... Teachers with higher salaries are concentrated in high income and low minority schools. Furthermore, pupil-teacher ratios are higher in schools with larger minority and low-income student populations.... Educational units with higher proportions of low-income and minority students are allocated fewer fiscal and educational resources than are more affluent educational units, despite the probability that these students have substantially greater need for both (p.425).

The situation has not improved in most states over the last decade, as recent lawsuits challenging inequality in Alabama, California, Louisiana, New York, and elsewhere demonstrate. In combination, policies associated with school funding, resource allocations, and tracking leave poor and minority students with fewer and lower quality books, curriculum materials, laboratories, and computers; significantly larger class sizes; less qualified and experienced teachers; and less access to high quality curriculum. The fact that the least qualified

teachers typically end up teaching the least advantaged students is particularly problematic, given recent studies that have found that teacher quality is one of the most important determinants of student achievement (for a review, see Darling-Hammond, 2000). Not only is teacher quality a significant prerequisite for learning, but low-income and minority students are least likely to receive well-qualified, highly-effective teachers (Ferguson, 1991; Oakes, 1995; Sanders and Rivers, 1996; NCES, 1997a). Some evidence suggests that differences in the quality of teachers available to poor and minority children may explain nearly as much of the variance in student achievement as socioeconomic status.

These problems can be covered up and exacerbated by grade retention policies that purport to provide accountability. Nationally, retention rates for low-income children are at least twice those for high-income students. Students who are retained in grade are disproportionately representative of racial and ethnic minorities and populations whose dominant language is other than English (Illinois Fair Schools Coalition, 1985; Medway, 1985; Shepard and Smith, 1986; Smith and Shepard, 1987; Walker and Madhere, 1987). Thus, the students who receive the scantiest resources, the least qualified teachers, the poorest physical facilities, and the most restricted access to quality learning opportunities are supposed to be "fixed" by being held back.

The Chicago study notes that the failure to invest in improved teaching is an unrecognized problem in the city's reform strategy, which has tried to rely on a highly-scripted centrally developed curriculum (which by design assumes that students learn in the same ways and at the same pace) and grade retention as its major tools: "Thus the administration has worked to raise test scores among low-performing students without having to address questions regarding the adequacy of instruction during the school day or spend resources to increase

teachers' capacity to teach and to meet students' needs more successfully" (Roderick et al., 1999, p. 57).

Where the failure to learn is a result of inadequate teaching and where the system's primary response is to require children to experience that inadequate teaching again, it is doubtful that such a policy increases the system's accountability to parents and students. Where a side-effect of the policy is that large numbers of students drop out of school, thus creating a societal burden of under-educated youth who are unable to function in the labor market and who increasingly join the welfare or criminal justice systems rather than the productive economy, the schools' accountability to the greater society is also reduced. Society as a whole does not benefit from school policies that claim to heighten accountability by pushing low achievers out of school to make test scores look better – a result that has been documented in several cities – or by failing to offer education that enables these students to learn.

Institutional responses to test-based incentives

Unfortunately, most cities and states have used test-based reform strategies that rely on cross-sectional measures of student scores for different populations of students (e.g. average scores for 8th graders in a given year are compared to average scores for a different group of 8th graders in the prior year), rather than longitudinal assessments of student gains for students who remained in a given school over a period of time. Because schools' average scores on any measure are sensitive to changes in the population of students taking the test, and such changes can be induced by manipulating admissions, dropouts, and pupil classifications, policies that use school test scores for allocating school sanctions have been found to result in several unintended negative consequences. These include labeling large numbers of low-scoring students for special education placements so that their scores won't "count" in school reports, retaining students in

grade so that their relative standing will look better on "grade-equivalent" scores, excluding low-scoring students from admission to "open enrollment" schools, and encouraging such students to leave schools or drop out (Shepard and Smith, 1988; Koretz, 1988; Smith et al., 1986; Darling-Hammond, 1991, 1992; Allington and McGill-Franzen, 1992).

This occurs because the policies create incentives for schools to keep out of the testing pool – or the school itself -- students who will lower the average scores. This includes children who are handicapped, limited English-speaking, or from educationally disadvantaged environments. Smith and colleagues explained the widespread engineering of student populations that he found in his study of New York City's implementation of performance standards as a basis for school level sanctions:

> (S)tudent selection provides the greatest leverage in the short-term accountability game....The easiest way to improve one's chances of winning is (1) to add some highly likely students and (2) to drop some unlikely students, while simply hanging on to those in the middle. School admissions is a central thread in the accountability fabric (Smith et al., 1986, pp. 30-31).

In some cases, policies that reward or punish schools for aggregate test scores have created a distorted view of accountability, in which beating the numbers by manipulating student placements overwhelms efforts to serve students' educational needs well. These policies may also further exacerbate existing incentives for talented staff to opt for school placements where students are easy to teach, and school stability is high. Capable staff are less likely to risk losing rewards or incurring sanctions by volunteering to teach where many students have special needs and performance standards will be more difficult to attain. This outcome was recently reported as a result of Florida's recent use of aggregate test scores, unadjusted for student characteristics, for school rewards and sanctions. Qualified teachers were leaving the schools rated D or F "in droves" according to news reports at the start of the 1999 school year (DeVise, 1999; Fischer,

1999), to be replaced by teachers without experience and often without training. As one principal queried, ``Is anybody going to want to dedicate their lives to a school that has already been labeled a failure?"

Ironically, this approach to accountability compromises even further the educational chances of disadvantaged students, who are already served by a disproportionate share of those teachers who are inexperienced, unprepared, and underqualified. This outcome would be further exacerbated by current proposals by President Bush to reduce federal funds to schools that have lower test scores. Critics have argued that applying sanctions to schools with lower test score performance penalizes already disadvantaged students twice over: having given them inadequate schools to begin with, society now punishes them again for failing to perform as well as other students attending schools with greater resources and more capacity. These kinds of sanctions discourage good schools from opening their doors to educationally needy students and place more emphasis on manipulating scores by eliminating or keeping out low-scoring students than on improving schools.

These outcomes have been noted of reforms in several states. For example, after the Regents Test reforms of the early 1980s in New York State, studies found evidence of schools retaining students and placing them in special education to increase average school performance (Allington and McGill-Franzen, 1992) and encouraging low-scoring secondary students to leave school entirely (Smith et al., 1986). By 1992, New York's graduation rates had dropped to only 62%, leaving the state ranked 45th in the country on this measure (Feistritzer, 1993).

Similarly, Atlanta, Georgia instituted a pupil progression policy in 1980 based on test score thresholds for each elementary grade. High failure rates and repeated retentions led to increased dropout rates. The high school completion rate in Atlanta dropped to 65% by 1982 and

to 61% by 1988. A 1988 state policy set up additional test thresholds for promotion and graduation. This policy exacerbated the declines in graduation in Atlanta and elsewhere across the state. As Gary Orfield and Carole Ashkinaze (1991) noted:

> Although most of the reforms were popular, the policymakers and educators simply ignored a large body of research showing that they would not produce academic gains and would increase dropout rates. In other words, this was a policy with no probable educational benefits and large costs. The benefits were political and the costs were borne by at-risk students. The damage was psychological as well as educational, increasing the likelihood that at-risk students would drop out before receiving their diplomas; school districts were also hurt by the diversion of resources to repetitive years of education for many students (139).

An analysis of the test-based reform strategies in Georgia and South Carolina, both of which tied rewards and sanctions to annual tests at each grade level found that neither state realized any gains in achievement on the National Assessment of Educational Progress during the 1990s, although both experienced declines in high school graduation rates (Darling-Hammond, 2000).

Recent analyses of test-based reforms instituted in Texas in the 1980s have pointed to these and other problems. Although ostensible gains in scores on the TAAS tests have caused the state to be hailed as the "Texas Miracle," a number of studies have found that the outcomes are less positive than they appear. First, studies by the Center for Research and Evaluation on Testing (Haney, 2000) and by the Intercultural Development Research Association (1986) have found that both retention rates in 9th grade and dropout or attrition rates for high school students have increased substantially since the late 1980s. Both studies found that fewer than 50% of African American and Latino 9th graders progress to 12th grade and graduation four years later, and only about 70% of white 9th graders reach graduation. Haney (2000) found evidence that a growing number of low-scoring students leave school as early as 8th or 9th grade, before their

scores are factored into school accountability rankings. The effects are most pronounced for minority students:

> In 1990-91, Black and Hispanic high school graduates relative to the number of Black and Hispanic students enrolled in grade 9 three years earlier fell to less than 0.50 and this ratio remained just about at or below this level from 1992 to 1999. (The corresponding ratio had been about 0.60 in the late 1970s and early 1980s).... From 1977 until about 1981 rates of grade 9 retention were similar for Black, Hispanic, and White students, but since about 1982, the rates at which Black and Hispanic students are denied promotion and required to repeat grade 9 have climbed steadily, such that by the late 1990s, nearly 30% of Black and Hispanic students were "failing" grade 9 and required to repeat that grade.

Haney's report and Texas Education Agency (TEA) analyses agree that dropout rates in Texas are substantially higher for students retained in 9th grade than for any other group, Although the TEA data provide a rosier picture of overall graduation rates, since they do not count as dropouts the large number of students who are transferred to GED programs and fail to finish them.

Several recent studies have produced empirical data that cast doubt on the gains noted on the state TAAS tests, observing that Texas students have not made comparable gains on national standardized tests or on the state's own college entrance test (Klein et al., 2000). These studies have variously suggested that teaching to the test may be raising scores on the state high-stakes test in ways that do not generalize to other tests that examine a broader set of higher order skills; that many students are excluded from the state tests to prop up average scores; and that the tests have been made easier over time to give the appearance of gains (Haney, 2000; Gordon and Reese, 1997; Hoffman et al., 1999; Klein et al., 2000; Stotsky, 1998).

For all of these reasons and others, the American Psychological Association, American Educational Research Association, and the National Council on Measurement in Education have issued standards for the use of tests that indicate that test scores are too limited and unstable a

measure to be used as the sole source of information for any major decision about student placement or promotion. A recent report of the National Research Council on high stakes testing concluded:

> Scores from large-scale assessments should never be the only sources of information used to make a promotion or retention decision.... Test scores should always be used in combination with other sources of information about student achievement (Heubert an Hauser, 1999, p. 286).

The test-based accountability systems in dozens of states and urban school systems stand in direct contravention to these professional standards. However, the negative effects of grade retention and graduation sanctions should not become an argument for social promotion -- that is, the practice of moving students through the system without ensuring that they acquire the skills that they need. What are the alternatives? There are at least four complementary strategies that have been shown to have substantial influences on student learning:

(1) enhancing preparation and professional development for teachers to ensure that they have the knowledge and skills they need to teach a wider range of students to meet the standards;

(2) redesigning school structures to support more intensive learning – including creating smaller school units (within an optimal size of 300-500) and schools that team teachers to work with smaller total numbers of students for longer periods of time;

(3) employing schoolwide and classroom performance assessments that support more coherent curriculum and better inform teaching; and

(4) ensuring that targeted supports and services are available for students when they are needed.

Some urban districts have used these strategies to upgrade student learning and to create genuine accountability to parents and students. These success stories offer a very different model for standards-based reform, one that rests on professional accountability and the use of standards and assessments as a stimulus for professional development and curricular reform rather than as punishments for schools and students. Three examples are offered here: the statewide reforms in Connecticut that have supported substantial improvements in a number of cities (featured here are New Britain, Norwalk, and Middletown – among the state's lowest-income and once lowest-achieving districts); New York City's School District #2, and New Haven, California.

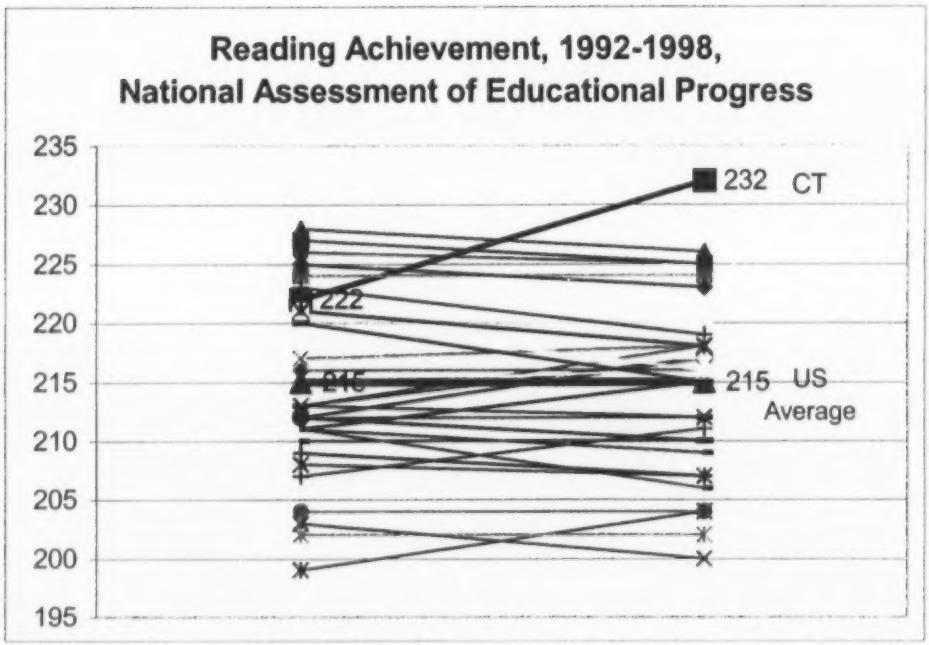
Connecticut

Connecticut provides an especially instructive example of how state level policy makers have used a standards-based starting point to upgrade teachers' knowledge and skills as a means of improving student learning. Since the early 1980s, the state and several of its urban centers, described below, have pursued a purposeful and comprehensive teaching quality agenda. The Connecticut case is a story of how bipartisan state policy makers implemented a coherent policy package over more than 15 years. They used teaching standards, followed later by student standards, to guide investments in school finance equalization, teacher salary increases tied to higher standards for teacher education and licensing, curriculum and assessment reforms, and a teacher support and assessment system that strengthened professional development.

Connecticut's teacher assessments and preparation requirements ensure that every entering teacher has strong content and pedagogical knowledge to enable him or her to teach a wide range of diverse learners well – including those who have special education needs and English language learning needs. Standards-based professional development opportunities have

dramatically upgraded the knowledge and skills of the veteran teaching population. Student assessments are aimed at higher order thinking and performance skills and are used to evaluate and continually improve practice. While the highly public reporting system places strong pressure on districts and schools to improve their practice, the student assessments are not used for rewards or punishments for students, teachers, or schools. Rather than pursue a single silver bullet or a punitive approach that creates dysfunctional responses, Connecticut has made ongoing investments in improving teaching and schooling through high standards and high supports.

Dramatic gains in student achievement (accompanied by increases rather than declines in student graduation rates) and a plentiful supply of well-qualified teachers are two major outcomes of this agenda. By 1998, Connecticut's 4th grade students ranked first in the nation in reading and mathematics on the National Assessment of Educational Progress (NAEP), despite increased student poverty and language diversity in the state's public schools during that decade (NCES, 1997b; NEGP, 1999). The proportion of Connecticut 8th graders scoring at or above proficient in reading was also first in the nation, and Connecticut was only the top performing state in writing, but the only one to perform significantly better than the U.S. average. A 1998 study linking the NAEP with the Third International Math and Science Study (TIMSS) found that, in the world, only top-ranked Singapore outscored Connecticut students in science (Baron, 1999). While there remains an achievement gap between white students and the large and growing minority student population, the more than 25% of Connecticut's students who are black or Hispanic substantially outperform their counterparts nationally as well (Baron, 1999).



In explaining Connecticut's reading achievement gains, a recent National Educational Goals Panel report (Baron, 1999) cited the state's teacher policies as a critical element, pointing to the 1986 Education Enhancement Act, as the linchpin of the teacher reforms. In this omnibus bill, Connecticut coupled major increases in teacher salaries with greater equalization in funding across districts, higher standards for teacher education and licensing, and substantial investments in beginning teacher mentoring and professional development. An initial investment of $300 million was used to boost minimum beginning teacher salaries in an equalizing fashion that made it possible for low-wealth districts to compete in the market for qualified teachers. The average teacher's salary increased from a 1986 average of $29,437 to a 1991 average of $47,823 (Fisk, 1999). These grants were provided on an equalizing basis to enable poor districts to better compete in the market for qualified teachers. Districts were given incentives to hire qualified

teachers because salary grants were calculated on the basis of fully certified teachers only, and emergency credentials were phased out.

To further ensure an adequate supply of qualified teachers, the state offered incentives including scholarships and forgivable loans to attract high-ability teacher candidates, especially in high-demand fields, and encouraged well-qualified teachers from other states to come to Connecticut through license transportability reforms. An analysis of the outcomes of this set of initiatives found that they eliminated teacher shortages, even in the cities, and created surpluses of teachers within three years of its passage (Connecticut State Department of Education, 1990). These surpluses have been maintained since, allowing districts – including urban school districts -- to be highly selective in their hiring and demanding in their expectations for teacher expertise.

At the same time, the state raised teacher education and licensing standards by requiring a major in the discipline to be taught plus extensive knowledge of teaching and learning as part of preparation (including knowledge for all teachers about literacy development and the teaching of special needs students); instituted performance-based examinations in subject matter and knowledge of teaching as a basis for receiving a license; created a state-funded beginning teacher mentoring program which supported trained mentors for beginning teachers in their first year on the job; and created a sophisticated assessment program using state-trained assessors for determining who could continue in teaching after the initial year.

Connecticut also required teachers to earn a master's degree in education for a continuing license and supported new professional development strategies in universities and school districts. Recently, the state has further extended its performance-based licensing system to incorporate the new INTASC standards and to develop portfolio assessments modeled on those of the National Board for Professional Teaching Standards. As part of ongoing teacher

education reforms, the state agency has supported the creation of 17 professional development schools linked to local universities and more than 100 school-university partnerships. In addition, Connecticut has developed courses on teacher and student standards that can be applied toward the required Master's degree. The state also funds and operates a set of Institutes for Teaching and Learning.

Connecticut's portfolio assessments for beginning teacher licensing are modeled on those of the National Board for Professional Teaching Standards; they examine directly whether a teacher is able to teach to Connecticut's student learning standards in each content area in which the teacher teachers. The performance assessments examine teacher plans, videotapes of lessons, student work, and teacher analyses of their practice. They are developed with the assistance of teachers, teacher educators, and administrators: Hundreds of educators are convened to provide feedback on drafts of the standards; and many more are involved in the assessments themselves, as cooperating teachers and school-based mentors who work with beginning teachers on developing their practice, as assessors who are trained to score the portfolios; and as expert teachers and teacher educators who convene regional support seminars to help candidates learn about the standards and the portfolio development process. Preparation is organized around the examination of cases and the development of evidence connected to the standards.

Together, these activities can have far-reaching effects. By one estimate, more than 40% of Connecticut's teachers have gone through the process as new teachers or have served as assessors, mentors, or cooperating teachers unde the earlier beginning teacher performance assessment or the new portfolios. By the year 2010, 80% of elementary teachers, and nearly as many secondary teachers, will have participated in the new assessment system as candidates, support providers, or assessors. Because the assessments focus on the development of teacher

competence, are tightly tied to student standards, and lead to sophisticated analysis of practice, the assessment system serves as a focal point for improving teaching and learning.

In addition to the state's major investments in teaching quality, the Goals Panel report also pointed to the thoughtful use of student standards and assessments in Connecticut. In 1987, following the teaching reforms, student learning standards were adopted in an early effort to link teacher education standards with expectations for teaching. In 1993-94, the student standards were updated to emphasize higher order thinking skills and performance abilities, and new assessments were developed; these are constructed response and performance assessments that measure reading and writing authentically and reflect more challenging learning goals than the previous tests.

Also critical is the fact that these assessments are precluded from being used for promotion or graduation of students. Instead, they are used for ongoing improvements in curriculum and teaching. The Goals Panel report noted the benefits of the state's low-stakes testing approach, which emphasize reporting and analysis strategies that support the wide dissemination of the standards and test objectives along with widespread professional development around literacy and the teaching of reading. The State Department of Education also supports the use of test results for educational improvement by giving districts computerized data that allow analyses at the district, school, teacher, and individual pupil level. The Department assists districts in analyzing the data in ways that permit diagnosis of needs and areas for concentrated work (Baron, 1999). The state then provides targeted resources to the neediest districts to help them improve, including funding for professional development for teachers and administrators, preschool and all-day kindergarten for students, and smaller pupil-teacher ratios, among other supports.

The Goals Panel study notes that this approach to assessment has enabled districts to clarify their teaching priorities and has helped galvanize district efforts to make major revisions and improvements in their reading instruction. At the same time, the targeted provision of resources to the state's neediest districts through categorical grants has enabled these districts to enhance their reading initiatives and to begin to close the gap between their scores and those statewide (Baron, 1999).

Among the ten Connecticut districts that made the greatest progress in reading between 1990 and 1998, three -- New Britain, Norwalk, and Middletown -- are urban school systems in the group identified as the state's "neediest" districts based on the percentage of students eligible for free lunch programs and their state test scores.

District	Grade Level	1993 CMT Index Score	1998 CMT Index Score	Gain in Average CMT Score
STATE AVERAGE	Grade 4	56.9	65.5	+8.6
	Grade 6	68.0	74.2	+6.2
	Grade 8	69.9	75.5	+5.6
Middletown	Grade 4	51.8	65.7	+13.9
	Grade 6	67.0	74.2	+7.2
	Grade 8	64.7	75.6	+10.9
Norwalk	Grade 4	46.6	58.6	+12.0
	Grade 6	55.3	62.7	+7.4
	Grade 8	53.8	66.4	+12.6
New Britain	Grade 4	36.3	47.4	+11.1
	Grade 6	35.0	45.6	+10.6
	Grade 8	38.5	52.3	+13.8

Follow up studies in these districts identified a number of state-level policies and related local strategies as contributing to this success (Baron, 1999). Among them were teacher policies that have enabled districts to hire and retain highly qualified teachers who had been prepared to teach a wide range of learners as well as the beginning teacher program that provided state-training for all mentors (required for every beginning teacher) and increased the knowledge and

skills of veteran teachers along with beginners involved with the program. In addition, district respondents described state- and locally-supported intensive professional development around the teaching of reading. Consistent with the Common Core of Learning and the state assessments, professional development funds were orchestrated to improve teachers' knowledge of how to teach reading through a balanced approach to whole language and skill-based instruction, how to address reading difficulties through specific intervention strategies, and how to diagnose and treat specific learning disabilities, which was addressed as part of the state's efforts to prepare all teachers to meet special education needs. Most of the districts had developed cadres of teacher trainers or coaches who were experts in literacy development and who were available to work with colleagues in the schools, offering demonstration teaching as well as classroom coaching. A number used State grants to sponsor multi-week literacy workshops held in the summer focused on the teaching of at-risk readers.

The approaches to reading instruction used in sharply improving districts rely on the enhanced teacher knowledge spurred in Connecticut's teacher education reforms and represented in the state's teaching assessments: systematic teaching of reading and spelling skills (including linguistics training that goes beyond basic phonemic awareness); use of authentic reading materials – children's literature, periodicals, and trade books – along with daily writing and discussion of ideas; ongoing assessment of students' reading proficiency through strategies like running records, miscue analyses, and analysis of reading, writing, and speaking samples; and intervention strategies for students with reading delays, such Reading Recovery, which was used in 9 of the 10 sharply improving districts and is widely used across the state (Baron, 1999).

District administrators noted the importance of the system's coherence in allowing them to pursue these sophisticated strategies for teaching and learning. In addition to their work on

teacher development, they described how they had realigned district curriculum and instruction to the student learning standards and assessments, and how they had used of the rich information about student performance made available by the CSDE as the basis for school problem solving and teachers' individual growth plans (the latter are part of the teacher evaluation system). They also credited as important the fact that the state assessments measured reading and writing in reasonably authentic ways, that the preparation and professional development programs were supportive of the same approaches, and the fact that beginning teachers were coming to them better and better prepared to teach to these standards using successful pedagogical strategies, while veterans also had many opportunities to develop.

The quality of teaching in Connecticut can be traced directly to the implementation of an increasingly well-developed statewide infrastructure that has been designed to encourage high-quality teaching by (a) linking salaries to high standards for preparing, entering, and remaining in teaching, (b) providing intensive support and assessment of beginning teachers, and (c) requiring and supporting continued high-quality professional development for teachers and administrators. These factors have helped establish a foundation of professional expertise that can ensure the success of other organizational policies and practices, such as analysis of student achievement results, linking school improvement plans and teacher evaluations to student achievement, and aligning expectations and assessments for students with high standards for teachers.

New York City District #2

A remarkably similar set of strategies has produced similar results in New York City's Community School District #2, an extremely diverse, multilingual district of 22,000 students of whom more than 70% are students of color and more than half are from families officially classified as having incomes below the poverty level.[1] More than 100 different languages are

spoken in the collective homes of District #2 students, a large share of whom are recent immigrants. During the decade-long tenure of superintendent Tony Alvarado, from 1987 to 1997, the district rose from 11th to 2nd in the City in student achievement in reading and mathematics, scoring above New York State norms as well as New York City averages, even while the population of the district grew more more language diverse.

Studies of District #2 have attributed these gains to the district's decision to make professional development the central focus of management and the core strategy for school improvement. The strong belief governing the district's efforts is that student learning will increase as the knowledge of educators grows (Elmore & Burney, 1997). Rather than treating professional development as a discrete function implemented with a set of disparate non-systemic activities, District 2 makes professional development around common standards of teaching the most important focus of all district efforts, its most prominent discretionary budgetary commitment, and a key part of every leader's and every teacher's job.

After consolidating categorical and other discretionary funds and focusing them on a coherent, curriculum-embedded program of professional learning, District 2 moved most of its central office personnel positions back to school sites to focus on the improvement of practice. In a set of moves intently focused on enhancing professional accountability, Alvarado aggressively recruited instructionally knowledgeable teachers and principals, created pointed expectations and opportunities for professional development around the deepening of instructional practice – first in literacy and then in mathematics – and replaced through retirements, "counseling out," and personnel actions those under-skilled principals and teachers who were unable or unwilling to develop their practice. He evaluated principals and school directors on their success in improving the quality of teaching in their buildings, assisting

principals as well as teachers in learning about best practices in teaching literacy and mathematics and then holding leaders accountable for evaluating the quality of practice in their buildings, establishing networks with other principals and with higher education institutions to recruit student teachers and new teachers, and working with district personnel to ease the transition of ineffective teachers out of the district and prevent the transfer of ineffective teachers into the district.

At the same time he was exercising influence over the appointment of staff, Alvarado created 17 "Option Schools," small alternative schools that reorganized instruction to focus on greater personalization and more performance-based assessments to guide teaching, while encouraging the redesign of other schools. These efforts made practices known to support achievement more widespread in the district, including more small schools, grouping practices that keep teachers and students together for more than one year through practices like "looping," schedules that allow collaborative planning and professional development for teachers within the school day, and more coherent, intellectually challenging curriculum supported by ongoing diagnostic and performance assessments of student learning.

School redesign was joined with professional development in a conscious strategy to improve both teachers' expertise and schools' ability to support in-depth teaching and learning. Well-known for his efforts to create restructured schools and schools of choice when he was previously superintendent in District #4, Alvarado found that the creation of new alternatives, while useful for the schools where dynamic educators coalesced, did not go far enough in building knowledge for better practice in all schools and classrooms. As he explained, "When I moved to District 2, I was determined to push beyond the District 4 strategy and to focus more

broadly on instructional improvement across the board, not just on the creation of alternative programs" (Elmore & Burney, 1997).

Staff development in District 2 differs sharply from the one-shot workshop that expects teachers to take generic ideas unconnected to their ongoing work and apply them in the classroom. Rather, the prevailing theory is that changes in instruction occur only when teachers receive continuous oversight and support focused on the practical details of what it means to teach effectively. The district's extensive professional development efforts, which have paid off in rapidly rising student achievement, include several vehicles for learning. Instructional consulting services allow expert teachers and consultants to work within schools with groups of teachers in sustained ways develop to particular strategies, such as literature-based reading instruction. Intervisitation and peer networks are designed to bring teachers and principals into contact with exemplary practices. The district budgets for 300 total days each year to provide the time for teachers and principals to visit and observe one another, to develop study groups, and to pair up for work together. Off-site training includes intensive summer institutes that focus on core teaching strategies and on learning about new standards, curriculum frameworks, and assessments. These are always linked to followup through consulting services and peer networks to develop practices further. The Professional Development Laboratory allows visiting teachers to spend 3 weeks in the classrooms of expert resident teachers who are engaged in practices they want to learn. Oversight and evaluation of principals focuses on their plans for instructional improvement in each content area, as does evaluation of teachers. There is close, careful scrutiny of teaching from the central office as well as the school and continual pressure and support to improve its quality. As Elmore and Burney (1997) explain:

> Shared expertise takes a number of forms in District 2. District staff regularly visit principals and teachers in schools and classrooms, both as part of a formal evaluation

> process and as part of an informal process of observation and advice. Within schools, principals and teachers routinely engage in grade-level and cross-grade conferences on curriculum and teaching. Across schools, principals and teachers regularly visit other schools and classrooms. At the district level, staff development consultants regularly work with teachers in their classrooms. Teachers regularly work with teachers in other schools for extended periods of supervised practice. Teams of principals and teachers regularly work on districtwide curriculum and staff development issues. Principals regularly meet in each others' schools and observe practice in those schools. Principals and teachers regularly visit schools and classrooms within and outside the district. And principals regularly work in pairs on common issues of instructional improvement in their schools. The underlying idea behind all these forms of interaction is that shared expertise is more likely to produce change than individuals working in isolation.

A key feature of these strategies is that they have focused intensely for multiple years on a few strands of content-focused training designed to have cumulative impact over the long term, rather than changing workshop topics every in-service day or picking new themes each year. The district has sponsored 8 years of intensive work on teaching strategies for literacy development and 4 years on mathematics teaching. District 2's approach began with reading and writing because this focus provided a readily available way for the district to demonstrate improvement in academic performance in an area that was important on city-wide assessment measures and because literacy was important in the context of the district's linguistic and ethnic diversity. New York City's development of more performance-oriented assessments in reading and mathematics in the early 1990s provided more useful targets for these instructional reforms.

As in Connecticut, Reading Recovery training for an ever-widening circle of teachers created the first foundations of the teacher development initiative. This effort was used to improve teachers' knowledge about how to teach reading to their entire classrooms of students, not just to provide one-on-one tutoring to students with special reading needs. Ongoing work focused on whole language approaches to the teaching of reading and writing, with integration of specific work on reading skills and strategies focused by individual student assessment through tools like the Primary Language Record that helped teachers develop documentation through

running records, miscue analyses, and analysis of student work samples. As district staff, consultants, and principals learned how to change teaching practice through the literacy initiative, drawing on local university supports like Teachers College's Writing Institute and the Lehman College Literacy Center as well as district expertise, Alvarado began a parallel effort in mathematics using a similar model that drew in part on mathematics coaches trained at Bank Street's School of Education.

Much of this work occurred within the context of changes in New York State's learning standards and curriculum frameworks that supported district efforts to develop more challenging, performance-oriented standards to be used in assessing student work. District #2, and later New York City, adopted the curriculum frameworks of the New Standards Project and formed an alliance with the University of Pittsburgh's new Institute for Learning, piloting its performance assessments of student learning which use portfolios and extensive student work samples as well as constructed response tests. Alvarado saw this emerging emphasis on standards as a logical extension of the District's efforts at instructional improvement. At the same time, he argued that introducing the standards and assessments before principals and teachers had had extensive experience with instructional improvement would have been a mistake. "You can kill a lot of the learning that you need in the system by insisting that it all has to line up with some item on a test," he explained. On the other hand, he felt that standards and assessment are logical extensions of an emphasis on professional development as a mechanism of instructional improvement (Elmore and Burney, 1997).

While assessments of student learning are a critical element in the overall improvement strategy, the incentive structures are explicitly aimed at improving professional accountability – that is, the capacity and commitment of educators to teach well – rather than hoping for

improved learning by increasing the amount of testing or adding sanctions to test scores. Elmore and Burney note, "Accountability within the system is expressed in terms of teachers' and principals' objectives for instructional improvement.... (M)anagement is operationally defined as helping teachers to do their work better, and work is defined in terms of teaching and learning." Professional accountability has meant high-stakes attached to hiring and retaining high-quality teachers and principals, rather than stakes that punish students who do not succeed. While Alvarado replaced 80% of the principals in District 2 in his first four years, about 50% of teachers in the district were replaced over the course of 8 years, not through random attrition but through careful recruitment and replacement. Elmore and Burney report:

> This attitude toward the centrality of personnel decisions has begun to permeate, in turn, principals' attitudes toward the hiring of teachers. Most of the principals we interviewed in the system said spontaneously, without any prompting, that the key determinant of their capacity to meet their school-level objectives was the quality of their teachers and that they had learned how to exercise more influence on the process of recruiting, hiring, nurturing, retaining, and firing, or counseling-out, of teachers in their schools.

The emphasis on professional accountability, while uncomfortable for those not interested in improving, also created a positive professional culture in the district. Elmore and Burney also note:

> Most principals and teachers with whom we spoke reported that they were gratified, energized, and generally enthusiastic, if sometimes a bit intimidated, by the attention they received through District 2's professional development strategy. They report attending professional development activities outside the district or conducting visits to other schools and districts and being impressed with the amount of attention that teaching and learning receive in District 2. Teachers from outside the district who attend District 2-sponsored summer professional development activities often report that they have heard that the district is the place to be if you are interested in good teaching, and they comment favorably on the range of professional development activities available to District 2 teachers and principals. Outsiders also comment on the (to them) unusual practice of principals attending content-centered professional development activities with teachers from their schools.

The District 2 case shows how a district can mobilize resources to support sustained improvement in teaching practice and substantial improvements in student learning. In addition to the highly focused strategies the District uses to improve the quality of teaching practice system-wide, there are targeted efforts for students who do not initially succeed. In addition to the use of Reading Recovery strategies, Alvarado made investments in teacher training to teach English language learners and in highly expert special education services, replacing the common practice of assigning special needs students to untrained paraprofessionals with a strategy of hiring highly trained special educators who work with students but also share their expertise with other teachers, so that "regular" education teachers, too, can become more expert. Rather than using widespread grade retention, Alvarado focused these services on students with lagging achievement and assigned students with the lowest scores to the most expert teachers, rather than the most inexperienced and least well trained teachers, as is the custom in most cities.

These practices have been continued in the years since, under the leadership of an interim superintendent who had been Alvarado's deputy and then a superintendent promoted from among the ranks of highly-able instructionally knowledgeable principals appointed during the early years of the reform. The combination of these efforts focused first on teaching standards and then on student standards over a period of more than a decade has developed a brand of accountability in which parents in District 2 – a growing number of whom are now returning from private schools – are assured that their students will be well-taught, not just much-tested.

New Haven, California

Another glimpse of the possible can be seen in the New Haven Unified School District, located midway between Oakland and San Jose, California, a district that serves approximately 14,000 students from Union City and south Hayward, ¾ of whom are students of color, most of

them low-income and working class.[2] Twenty years ago, the district was the lowest wealth district in a low wealth county, and it had a reputation to match. Families who could manage to do so sent their children elsewhere to school. Today, New Haven Unified School District, while still a low-wealth district, has a well-deserved reputation for excellent schools. Every one of its 10 schools has been designated a California Distinguished School, and schools at all three levels have been designated as exemplary by the US Department of Education. All have student achievement levels well above California norms and even further above the norms for similar schools. The district has had to close its doors to out-of-district transfers because schools are bulging at the seams. Still, families try every trick in the book to establish a New Haven district address because they know their children will be well taught (Snyder, 1999).

One key element of New Haven's success was its commitment, twenty years ago, to high standards for teachers. Like Tony Alvarado in District #2, when superintendent Guy Emanuele first entered his post in the early 1980s, he started by establishing high expectations for teachers. He recalls:

> The presence of... teachers who did not perform to high standards lowered academic achievement of students and ultimately led to lower morale among other teachers.... One of my first acts as superintendent was to tighten the teacher evaluation process and implement procedures that allowed for due process while still enabling the district to remove teachers who simply were not able or willing to address deficiencies in their performance. A concerted focus upon teacher evaluations resulted in a number of resignations. Now, with performance standards in place and clear expectations as to the need to exceed them, teachers respect the district's effort to maintain high instructional standards, and rarely is a teacher terminated. Furthermore, the district's reputation in this regard draws high-achieving teachers, deters those who are not as committed, and generally elevates the status of the teaching profession (quoted in Snyder, 1999).

The district held administrators accountable for assessing teachers and providing necessary supports for teachers to meet expectations. New Haven put together very thorough evaluation procedures requiring the systematic collection of data -- no more "drive-by" teacher

observations. The responsibility for assuring the caliber of all teachers in all schools was a powerful incentive for making good initial hires. Making good hires required the district to revamp its recruitment and retention strategies to guarantee that qualified candidates would know about, wish to come to, and want to stay in the New Haven district.

Thirty years ago New Haven did what many districts continue to do today: Wait until the last minute and see what is available in the way of teachers. New Haven learned that even in a buyer's market, this is a short-sighted approach. They began to seek out exceptional teachers, simplify the application process, make decisions, and offer contracts in a manner that was timely and respectful of candidates. Over time, the district built support systems and teaching conditions that would retain exceptional teachers, and eventually it became involved in strong partnerships for preservice teacher education. Today, the district can afford to be choosy, recruiting with an eye toward teachers with the skills and dispositions to grow within the teacher learning environments the district supports. Unlike many other urban districts with similar student populations, New Haven does not have recruitment crises annually because of the low attrition rate of its new and experienced teachers (Snyder, 1999).

While school districts across California have scrambled in recent years to hire qualified teachers, New Haven had in place an aggressive recruitment system and a high quality training program with local universities that allowed it to continue its long-term habit of hiring universally well-prepared, committed, and diverse teachers to staff its schools. One factor in this success is that, despite its lower per pupil expenditures than many surrounding districts, New Haven spends the lion's share of its budget on teachers' salaries and then aggressively recruits and works to retain highly qualified teachers. For example, while nearby Oakland spends substantially more money per pupil, New Haven's beginning teacher salaries are nearly one-third

higher. In 1997-98, salaries in New Haven ranged from $37,604 to $70,373 -- the highest in the Bay Area and in the state's upper echelon -- despite New Haven's historic standing as one of the lowest-wealth districts in the state and the county (Snyder, 1999).

New Haven's total spending per ADA (a measure of per-pupil expenditure) was at that time $4,103, approximately the fifth percentile in the state and $2,337 per student below the highest per-pupil expenditure in the county. New Haven is not a rich district, but it affords quality because it:

- has flattened the traditional hierarchy of district and school bureaucracies (with 771 teachers and 50 "managers," nearly 94% of certified personnel work with children);
- allocates resources, including technology, to support and build teaching capacity; and
- creates multiple hybrid professional roles that enrich teacher learning while enhancing district policy and practice.

Rather than spending money on an array of special programs to address the problems created by inadequate teaching, the district decided to create a cadre of well-paid and highly-qualified teachers to avoid such problems in the first place.

A key to this strategy is coupling high salaries with high standards. Thus, while nearby Oakland hires large numbers of unqualified teachers while its personnel system keeps many qualified teachers from entering the system,[3] New Haven's personnel office uses technology and a wide range of teacher supports to recruit from a national pool of exceptional teachers.

Its web site posts all vacancies and draws inquiries from around the country. Each inquiry receives an immediate e-mail response. With the use of electronic information transfer (for example, the personnel office can send vacancy information directly to candidates and applicant files to the desktop of any administrator electronically), the district can provide information to people urban districts might never think would be available to them. Viable

applicants are interviewed immediately in person or via video-conference (through a local Kinko's), and it they are well-qualified with strong references, they may be offered a job that same day. Despite the horror stories one often hears about the difficulty of out-of-state teachers earning a California teaching credential, New Haven's credential analyst in the personnel office has yet to lose a teacher recruited from out-of-state in the state's credentialing maze.

Among the many factors contributing to the district's success in recruiting teachers and serving students, one significant strategy is the district's long-term investments in teacher education. The district was one of the first in the state to implement a Beginning Teacher Support and Assessment Program that provides support for teachers in their first two years in the classroom. All beginning teachers receive classroom support from a trained mentor who has released time for this purpose. The program is standards-based, using the California Standards for the Teaching Profession (CSTP) as the basis for both support and assessment. As in Connecticut's beginning teacher program, these standards point the attention of both beginning teachers and veteran teachers and principals who serve as members of their support teams to critical aspects of teaching, including effective strategies for diagnosing learning, planning curriculum to meet the needs of diverse learners, organizing and implementing instruction. Beginning teachers are guided by an individual induction plan developed with their support team, and they develop a portfolio that documents their growth toward the plan's goals. This is supplemented by a series of formal observations by support team members that guide additional goal-setting and a final assessment conducted in an interview format with the support team.

Many beginning teachers report that they chose to teach in New Haven because of the availability of this strong support for their initial years in the profession. In addition, in collaboration with California State University, Hayward, the district designed an innovative teacher education partnership that combines college coursework and an intensive internship

conducted under the close supervision of school-based educators. This program is guided by the same teaching standards as the beginning teacher program described above, creating coherence in teachers' pathways into teaching. Because interns function as student teachers who work in the classrooms of master teachers, rather than as independent teachers of record, the program simultaneously educates teachers while protecting students and providing quality education.

Throughout their careers teachers have access to a wide range of professional development opportunities throughout the year and, especially, in intensive summer work. For example, during the summer of 1997, approximately 65% of the district's teachers participated in district-sponsored staff development activities. The district has organized school schedules so that all teachers have the time to meet for 90 minutes each week to plan collaboratively. In addition, all of the professional work of the district engages teachers, thereby building and sharing their expertise and creating ownership in district reforms. In New Haven, classroom teachers enact the beginning teacher support and assessment program; develop curriculum; design technological supports; and create student standards and assessments.

As in District #2, standards for students have been developed and enacted as a professional development activity, using state and national frameworks as the starting point for engaging teachers in thinking through what students should know and be able to do, how it should be assessed, and what curriculum and instructional strategies could allow them to succeed. For example, using a combination of release time, after-school workshops, and extensive summer institutes, the district involved more than 100 teachers (nearly 40% of its K-4 teachers) in its language arts and mathematics standards committees during the 1996-97 year. New Haven began with this teacher-developed district-wide, comprehensive K-4 standards and assessment system that has since served as a prototype for all grade levels. This system consists of:

- clearly articulated performance standards with clear descriptions of seven different performance levels (from pre-readiness through independent) tied to grade-level expectations;
- a criterion-based parent reporting system for all K-4 students, including special education and second language learners;
- a three-strand assessment structure; and
- a database system that pulls together assessment, demographic, and intervention information for analysis and use in program planning and targeting student assistance.

The model is one of the few comprehensive standards systems in the country to incorporate a learner-centered developmental perspective with the more traditional accountability features of standards-setting efforts. The key to the standards and assessment system is not the testing itself but the web of supports activated by the assessments. The most fundamental use of the standards and assessment system is as a tool for classroom-level instructional planning. For example, in August each teacher receives a printout of the levels of each of his or her students' performance in reading, writing, and mathematics. Teachers initially use this information to design guided reading groups, target computer software, and assign home reading levels. Ongoing authentic assessment (for example, running records of reading) against the standards helps teachers continually modify these groupings. In addition, teachers use this assessment information to identify students needing tutoring during the after-school, extended-day program and/or homework support. On a more personal level, the database also helps maximize the match of primary-age students and intermediate-age reading buddies. And at the school level, educators use the system to guide changes in just about every educational arena, including staffing, instructional programming, resource allocation, and configuring classes (Snyder, 1999).

Such a program puts a major responsibility on teachers. What children know and are able to do must be clearly documented using students' classroom work, formal and informal assessment data, and teacher observation. This requires more than presenting information; it involves an expectation that the content of the standards be accessible to, and learned by, students at all performance levels. The purpose is not to label a child, but to develop a program that facilitates that child's development. Thus, standards and assessments are used to support the existing professional accountability structure by providing more information to guide collective as well as individual teaching practice. The fruits of these combined efforts to enact high standards for teaching and for learning show in New Haven's steadily increasing student achievement as well as its success in finding and keeping good teachers.

Improving the Chances of Student Success

Ultimately, accountability is not only about measuring student learning but actually improving it. Consequently, genuine accountability involves supporting changes in teaching and schooling that can heighten the probability that students meet standards. Unless urban school districts undertake systemic reforms in how they hire, retain, prepare, and support teachers and develop high quality teaching, the chances that all students will have the chance to meet new high standards are slight. There are at least three major areas where attention is needed:

(1) Ensuring that teachers have the knowledge and skills they need to teach to the standards;

(2) Providing school structures that support high quality teaching and learning; and

(3) Creating processes for school assessment that can evaluate students' opportunities to learn and can leverage continuous change and improvement.

Building Professional Capacity. The changes in teaching and assessment strategies needed to achieve new content and performance standards call for increased knowledge and

skills on the part of teachers. Teachers need deep understanding of subject matter, student learning approaches, and diverse teaching strategies to develop practices that will allow students to reach these new standards. To provide this kind of expertise to students, districts must pay much greater attention to the ways in which they recruit, hire, and support new teachers and the ways in which they support veteran teachers. Cumbersome and counterproductive personnel practices in many large district bureaucracies have resulted in the hiring of hundreds of untrained teachers when qualified personnel were available and in the attrition of far too many beginning teachers who are left to sink or swim without support. These practices create a continuous revolving door of inexperienced and under-prepared teachers in schools where student failure rates are the highest. Neither standards nor assessments will help students learn more effectively if they do not have a stable community of competent teachers to support them in their learning.

Until school systems address the dramatic inequalities in students' access to qualified teachers, other curriculum and assessment policies will prove ineffective in increasing achievement. In addition, schools and districts need to provide systematic supports for ongoing teacher learning in the form of time for shared teacher planning, opportunities for assessing teaching and learning, more exposure to technical expertise and resources, and opportunities for networking with other colleagues. These investments in building the capacities of teachers pay off in improved student outcomes (National Commission on Teaching and America's Future, 1996). In addition, as teachers learn to develop and use performance assessments, they discover more about their students and the effects of their teaching. This allows them to build more responsive and supportive teaching strategies that support the attainment of higher standards for a greater range of students (Darling-Hammond, Ancess, & Falk, 1995).

Providing these opportunities will require a clearer focus on teacher learning as a critical ingredient for enhanced student learning and as the most important preventive for the escalating costs of compensatory education, special education, grade retention, and other manifestations of student and school failure. Allocating resources to support teacher learning includes restructuring school time and staffing patterns to allow teachers time to work and learn together.

Structuring Schools to Support Student and Teacher Learning. As noted earlier, learning arrangements in which students work with the same teachers for more than one year facilitate higher levels of learning. In most high-achieving European and Asian countries, students stay with the same teacher for at least two years, and sometimes three or more. U.S. research has also found that smaller schools and schools that personalize instruction by keeping the same teachers with the same students for extended periods of time are associated with increased student achievement, more positive feelings toward self and school, and more positive behavior (NIE, 1977; Gottfredson & Daiger, 1979; Wehlage et al., 1989). Teachers are more effective when they know students well, when they understand how their students learn, and when they have more time with students to accomplish their goals.

Schools that have restructured to provide more shared planning and professional development time for teachers are also more successful at meeting the needs of diverse learners. When teachers can share knowledge with each other and can access expertise beyond the school, they learn how to succeed with students who require special insights and strategies. This kind of restructuring of time often requires rethinking staffing arrangements as well as schedules. In U.S. schools, where only 43% of total education staff are classroom teachers (as compared to 60-80% in many European schools and in Japan, for example), the costs of supporting non-teaching staff absorb the resources needed to provide planning time for teachers. Thus, whereas teachers

in many other countries have as much as 15 to 20 hours per week for joint planning and learning, U.S. teachers have only 3 to 5 hours weekly for class preparation, usually spent alone (National Commission on Teaching and America's Future, 1996). Creating time for teachers to work together often means reducing the number of nonteaching staff, pullout teachers, and specialists and reassigning them to teaching teams in order to increase person-power and time for classroom teaching.

Ensuring Opportunities to Learn. When students are to be held to the same set of learning standards, there must be means to ensure that all students have access to the conditions and resources needed for them to be able to meet these standards. Differential access to the resources that enable students' learning -- qualified teachers, adequate facilities, and high-quality materials – greatly impacts student achievement, disadvantaging those from under-resourced communities.

Along with standards for student learning, school systems should develop "opportunity-to-learn" standards -- standards for delivery systems and standards of practice -- to identify how well schools are doing in providing students with the conditions they need to achieve and to trigger corrective actions from the state and district. As Jeannie Oakes (1989) argues, information about resources and school practices is essential "if (policy makers) want monitoring and accountability systems to mirror the condition of education accurately or to be useful for making improvements" (p. 182). Those who would attempt to use standards in the quest for accountability and improvement can themselves be held accountable for making sound decisions only if they address questions of why outcomes appear as they do and make necessary changes in the conditions that influence learning.

The issue of standards and accountability cannot be separated from issues of teaching, assessment, school organization, professional development, and funding. Efforts aimed at better

supporting learning for all students so that they can successfully progress through school must include changes that address the overall fabric of education.

Academic success for a greater range of students will be facilitated by initiatives that:

- Use standards and authentic assessments of student achievement as indicators of progress for improved teaching and needed supports, not as arbiters of rewards and sanctions.
- Provide professional learning opportunities for teachers that build their capacity to teach ways that are congruent with contemporary understandings about learning, use sophisticated assessments to inform teaching, and meet differing needs.
- Encourage the design of classroom and grouping structures that create extended, intensive teacher-student relationships.
- Create strategies for school accountability that examine the appropriateness and adequacy of students' learning opportunities and create levers and supports for school change.

Ultimately, raising standards for students so that they learn what they need to know requires raising standards for the system, so that it provides the kinds of teaching and school settings students need in order to learn. Test-based grade retention and denial of diplomas as the major solutions to low achievement are merely a symbol of the failure of the system to teach successfully. Given the effects of these policies, such a strategy for accountability foreshadows the system's greater failure in the years ahead. Genuine accountability requires instead both higher standards and greater supports for student, teacher, and school learning.

---

1 This section draws from Richard Elmore and Deanna Burney (1997). Investing in teacher learning: Staff development and instructional improvement in Community School District #2, New York City. NY: National Commission on Teaching and America's Future.

2 This section draws from Jon Snyder, New Haven Unified School District: A Teaching Quality System for Excellence and Equity. NY: National Commission on Teaching and America's Future, Teachers College, Columbia University, 1999.

3 For a vivid illustration of the problem, see the John Merrow Report featuring Oakland and nearby New Haven, California, Teacher Shortage: False Alarm? NY: Merrow Report.

References

Allington, R. L. and McGill-Franzen, A. (1992). Unintended Effects of Educational Reform in New York, Educational Policy, 6 (4): 397-414.

Baron, J. B. (1999). Exploring High and Improving Reading Achievement in Connecticut. Washington: National Educational Goals Panel.

Baenen, N. (1988, April). A perspective after five years: Has grade retention passed or failed? Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Berne, R. (1995). Educational input and outcome inequities in New York State. In Robert Berne and Lawrence O. Picus (eds.), Outcome Equity in Education, pp. 191-223. Thousand Oaks, CA.: Corwin Press.

Carnegie Council on Adolescent Development, (1989). Turning Points: Preparing American Youth for the 21st Century. Task Force on Education of Young Addolescents, Washington, D.C.

Connecticut State Department of Education (1990). Impact of Education Enhancement Act, Research Bulletin, School Year 1990, No. 1.

Darling-Hammond, Linda (1991). The Implications of Testing Policy for Quality and Equality, Phi Delta Kappan, November 1991: 220-225.

Darling-Hammond, Linda (1992). Educational Indicators and Enlightened Policy. Educational Policy, 6 (3): 235-265.

Darling-Hammond, L. (1997). The right to learn. San Francisco: Jossey Bass.

Darling-Hammond, L. (2000). Teacher quality and student achievement. Educational Policy Analysis Archive, 8 (1): http://epaa.asu.edu/epaa/v8n1.

Darling-Hammond, L., Ancess, J., and Falk, B. (1995). Authentic Assessment in Action. New York: Teachers College Press.

Eads, G. (1990). Kindergarten Retention and Alternative Kindergarten Programs: Report to the Virginia Board of Education. ERIC ED 320.

Elmore, R. and Burney, D. (1997). Investing in teacher learning: Staff development and instructional improvement in Community School District #2, New York City. NY: National Commission on Teaching and America's Future.

Feistritzer, C.E. (1993). Report Card on American Education: A State-by-State Analysis, 1972-73 to 1992-93. Washington, D.C.: National Center on Education Information.

Ferguson, R., F. (1991). Paying for Public Education: New Evidence on How and Why Money Matters. Harvard Journal on Legislation, pp. 465-98.

Fisk, C. W. (1999). The emergence of bureaucratic entrepreneurship in a state education agency: A case study of Connecticut's education reform initiatives. Dissertation, University of Massachusetts-Amherst.

Gampert, R., and Opperman, P. (1988, April). Longitudinal Study of the 1982-83 "Promotional Gates Students". Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Gottfredson, G., and Daiger, D. (1979). Disruption in 600 Schools.
Baltimore: center for Social Organization of Schools, John Hopkins University.

Gordon, S.P. and Reese, M. (1997, July). High stakes testing. Worth the price? Journal of School Leadership, 7: 345-368.

Haney, W. (2000). The myth of the Texas miracle in education. Education Policy Analysis Archives, 8 (41): http://epaa.asu.edu/epaa/v8n41/

Hartocollis, A. (1999). Most assigned to summer school will not be promoted. New York Times, September 2, 1999.

Hess, A. (1986). Educational Triage in an Urban School Setting. Metropolitan Education, 2, pp.39-52.

Hess, G.A., Ells, E., Prindle, C., Liffman, P., and Kaplan, B. (1987). Where's room 185? How schools can reduce their dropout problem. Education and Urban Society, 19 (3): 330-55.

Heubert, J. and Hauser, R. (eds.) (1999). High stakes: Testing for tracking, promotion, and graduation. A report of the National Research Council. Washington, D.C.: National Academy Press.

Hoffman, J. V., Assaf, L., Pennington, J., & Paris, S. G. (in press). High stakes testing in reading: Today in Texas, tomorrow? The Reading Teacher.

Holmes, T., and Matthews, K. (1984). The Effects of Nonpromotion on Elementary and Junior High Pupils: A Meta-Analysis. Review of Educational Research, vol. 54, pp. 225-36.

Illinois Fair Schools Coalition, (1985). Holding Students Back: An Expensive Reform that Doesn't Work. Chicago: Author

Intercultural Development Research Association (1986, October). Texas school survey project: A summary of findings. San Antonio, TX: Intercultural Development Research Association.

Klein, S.P., Hamilton, L.S., McCaffrey, D.F., & Stetcher, B.M. (2000). What do test scores in Texas tell us? Santa Monica: The RAND Corporation.

Koretz, Daniel (1988). Arriving in Lake Wobegon: Are Standardized Tests Exaggerating Achievement and Distorting Instruction? American Educator, 12 (2), pp. 8-15, 46-52.

Koretz, D., & Barron, S. I. (1998). The validity of gains on the Kentucky Instructional Results Information System (KIRIS). Santa Monica, CA: RAND, MR-1014-EDU.

Koretz, D., Linn, R. L., Dunbar, S. B., & Shepard, L. A. (1991, April). The effects of high-stakes testing: Preliminary evidence about generalization across tests, in R. L. Linn (chair), The Effects of High Stakes Testing. Symposium presented at the annual meetings of the American Educational Research Association and the National Council on Measurement in Education, Chicago.

Labaree, D.F.. (1984). Setting the standard: Alternative policies for student promotion. Harvard Educational Review, 54 (1): 67-87.

Linn, R. L. (2000). Assessments and accountability. Educational Researcher, 29 (2), 4-16.

Linn, R. L., Graue, M. E., & Sanders, N. M. (1990). Comparing state and district test results to national norms: The validity of claims that "everyone is above average." Educational Measurement: Issues and Practice, 9, 5-14.

MacPhail-Wilcox, B. & King, R.A. (1986). Resource allocation studies: Implications for school improvement and school finance research. Journal of Education Finance, 11, pp.416-432.

Massachusetts Advocacy Center and the Center for Early Adolescence. (1988.) Before it's too late: Dropout Prevention in the Middle Grades. Boston: Author.

Medway, F. (1985). To promote or not promote? *Principal, 64* (3): 22-25.

Meisels, S. (1992, June). Doing harm by doing good: Iatrogenic effects of early childhood enrollment and promotion policies. *Early Childhood Research Quarterly 7* (2): 155-74.

Natale, J., A. (1991). Promotion or Retention? Ideas Are Changing – Again. Executive Educator, vol. 13, pp.-15-18.

National Commission on Teaching and America's Future (NCTAF) (1996). What Matters Most: Teaching for America's Future. New York: Author.

National Center for Education Statistics (NCES) (1997a). Digest of education statistics, 1997. Washington, DC: US Department of Education.

National Center for Education Statistics (NCES) (1997b). NAEP 1996 mathematics report card for the nation and the states. Washington, DC: US Department of Education.

National Education Goals Panel (NEGP) (1999). Reading achievement state by state, 1999. Washington, DC: U.S. Government Printing Office.

National Institute of Education (NIE). Violent Schools - Safe Schools: The Safe School Study Report to Congress. Washington, D.C.: Author, 1977.

Oakes, J. (1989). What Educational Indicators?: The Case of Assessing the School Context. Educational Evaluation and Policy Analysis, vol. 11, p. 182.

Oakes, Jeannie (1985). Keeping track: How schools structure inequality. New Haven: Yale University Press.

Oakes, J., and Lipton, M. (1990). Making the Best of Schools: A Handbook for Parents, teachers, and Policymakers. New Haven, CT: Yale University Press.

O'Day, Jennifer A. and Marshall S. Smith (1993). Systemic School Reform and Educational Opportunity, in Susan Fuhrman (ed.), Designing Coherent Education Policy: Improving the System. San Francisco: Jossey-Bass.

Orfield, G., and Ashkinaze C. (1991). The Closing Door: Conservative Policy and Black Opportunity. Chicago: University of Chicago Press, p. 139.

Ostrowski, P. (1987, November). Twice in one grade - A false solution. A review of the pedagogical practice of grade retention in elementary schools: What do we know? Should the practice continue? ERIC # ED300119

Price, Janet, Sara Schwabacher, and Ted Chittenden (1992). Report on the Multiple Forms of Evidence Study. NY: Fund for New York City Public Education.

Reese, C. M., Miller, K. E., Mazzeo, J., & Dossey, J. A. (1997). NAEP 1996 report card for the nation and the states. Washington, DC: National Center for Education Statistics.

Resnick, L. (1987). Education and learning to think. Washington, DC: National Academy Press.

Roderick, M., Nagaoka, J., Bacon, J., and Easton, J. (2000). Update: Ending Social Promotion: Passing, Retention, and Achievement Trends among Promoted and Retained Students, 1995-1999. Chicago: Consortium on Chicago School Research.

Roderick, M., Bryk, A.S., Jacob, B.A., Easton, J.Q. & Allensworth, E. (1999). Ending Social Promotion: Results from the first two years. Chicago: Consortium on Chicago School Research.

Safer, D. (1986). The stress of secondary school for vulnerable students. Journal of Youth and Adolescence, 15 (5): 405-17.

Sanders, W.L. & Rivers, J.C. (1996). Cumulative and Residual Effects of Teachers on Future Student Academic Achievement. Knoxville: University of Tennessee Value-Added Research and Development Center.

Shephard, L., and Smith, M.L. (1986). Synthesis of Research on School Readiness and Kindergarten Retention. Educational Leadership, 44(3) p. 86.

Shephard, L., and Smith, M.L. (1988). Flunking Kindergarten: Escalating Curriculum Leaves Many Behind. American Educator, 12(2): 34-38.

Smith, Frank, at al.(1986). High School Admission and the Improvement of Schooling. NY: New York City Board of Education.

Smith, M.L. and Shepard, L. (1987). "What Doesn't Work: Explaining Policies of Retention in the Early Grades. Phi Delta Kappan, 69(2): 123-124.

Snyder, J. (1999). New Haven Unified School District: A Teaching Quality System for Excellence and Equity. National Commission on Teaching and America's Future, NY: Teachers College Columbia University.

Stecher, B. M., Barron, S., Kaganoff, T., & Goodwin, J. (1998). The effects of standards-based assessment on classroom practices: Results of the 1996-97 RAND Survey of Kentucky Teachers of Mathematics and Writing (CSE Technical Report 482). Los Angeles: Center for Research on Evaluation, Standards, and Student Testing.

Stotsky, S. (1998). Analysis of Texas reading tests, grades 4, 8, and 10, 1995-1998. Report prepared for the Tax Research Association. http://www.educationnews.org/analysis_of_the_texas_reading_te.htm (August 1, 2000).

Walker. E., and Madhere, S. (1987). Multiple Retentions: Some Consequences for the Cognitive and Affective Maturation of Minority Elementary Students. Urban Education, 22, pp. 85-89.

Wasserman, J. (1999). 21,000 kids left back: Record number to repeat; social promotion ends. New York Daily News, September 2, 1999.

Wehlage, G. et al., (1989). Reducing the Risk: Schools as Communities of Support. New York: Falmer Press.

Wehlage, G.G., Rutter, R.A., Smith, G.A. Lesko, N., and Fernandez, R.R. (1990). *Reducing the risk: Schools as communities of support*. New York: Falmer Press.